Data in pragmatic research

Jucker, Andreas H

Abstract: This introductory chapter gives a broad-brush overview of the various types of data in the field of empirical research in pragmatics. It starts with a discussion of the various types of analytical units in pragmatics, taking as its starting point single utterances, which are contrasted to smaller units, such as deictic elements, stance markers, discourse markers, hedges and the like, as well as to larger units, such as sequences of utterances and entire discourses. Data for pragmatic research comes in different modalities. Spoken language and written language are the most obvious modalities, but digital language with its own complexities, sign language and non-verbal behaviour have recently become increasingly important as data for pragmatic research. Moreover, research data can be categorised on the basis of their location on four scalar dimensions. The first dimension concerns the amount of constraints on the interactants and the allowable contributions. The second dimension scales the level of fictionality or factuality of the language under observation. The third dimension assesses the amount of research interference in the production of the data, and the fourth dimension, finally, situates data according to the researcher focus between the two poles of small amounts of highly contextualized data to big data searches of largely decontextualized phenomena.

DOI: https://doi.org/10.1515/9783110424928-001

Originally published at:
Jucker, Andreas H (2018). Data in pragmatic research. In: Jucker, Andreas H; Schneider, Klaus P; Bublitz, Wolfram. Methods in Pragmatics. Berlin: De Gruyter, 3-36.
DOI: https://doi.org/10.1515/9783110424928-001

1. Data in pragmatic research

Andreas H. Jucker

Abstract: This introductory chapter gives a broad-brush overview of the various types of data in the field of empirical research in pragmatics. It starts with a discussion of the various types of analytical units in pragmatics, taking as its starting point single utterances, which are contrasted to smaller units, such as deictic elements, stance markers, discourse markers, hedges and the like, as well as to larger units, such as sequences of utterances and entire discourses. Data for pragmatic research comes in different modalities. Spoken language and written language are the most obvious modalities, but digital language with its own complexities, sign language and non-verbal behaviour have recently become increasingly important as data for pragmatic research. Moreover, research data can be categorised on the basis of their location on four scalar dimensions. The first dimension concerns the amount of constraints on the interactants and the allowable contributions. The second dimension scales the level of fictionality or factuality of the language under observation. The third dimension assesses the amount of research interference in the production of the data, and the fourth dimension, finally, situates data according to the researcher focus between the two poles of small amounts of highly contextualized data to big data searches of largely decontextualized phenomena.

1. Introduction

There is no research in pragmatics without data. Data – in one form or another – form the essence of what pragmatic research is about. Research – at a very basic level – consists in the search for generalizable patterns in the data. This is true for large computer searchable corpora, it is true for transcriptions of multi-party conversations and it is also true for thought experiments. Thus the researcher must start by collecting data in order to answer a specific research question. The type of data and the method of collecting the data are closely connected to the research question that drives the analysis and to the theoretical framework within which the analysis is carried out. A certain method of data collection will typically provide a very specific type of data and lend itself to a specific way of analysing it, or – viewed from the opposite direction – a certain research question will require a specific set of data that needs to be collected and analysed with a specific method. In general, we can distinguish four different aspects of research: 1) type of data, 2) method of data collection, 3) analysis of data and 4) theoretical framework.

https://doi.org/10.1515/9783110424928-001
In: A. H. Jucker, K. P. Schneider and W. Bublitz (eds.). (2018). *Methods in Pragmatics*, 3–36. Berlin/Boston: De Gruyter Mouton.

This opening chapter will focus on the first of them, and the following chapter by Schneider will be devoted to the second.

Often, researchers justify and defend a particular analytical and theoretical framework while they take a certain type of data and a certain method of data collection for granted. A certain type of data and a certain method of data collection are regularly presented as the only viable option. Unbiased overviews of different data collection methods and discussions of their inherent strengths and weaknesses are relatively rare (but see, for instance, Kasper and Dahl 1991; Kasper 2000; Jucker 2009a; Leech 2014: chapter 9). This volume starts from the premise that there is no single best type of data and no single best method for collecting data for pragmatic analyses. In fact, all four aspects of research mentioned above have to be assessed in relation to specific research questions (Jucker 2009a; Golato 2017: 21).

In general, the data of any pragmatic research is language used in actual contexts, and language is ever pervasive. We interact with other people, we watch television, attend lectures, read newspapers, consult user manuals, recite poems, surf the Internet, interact via social media, look at advertising messages, listen to public announcements on the train and so on and so forth. For each and every one of us the mix of communicative situations that we encounter every day is different, but every one of us is embedded in a flow of language. Even in our thoughts and in our dreams language plays an important role. Potentially all these situations, all these instances of language use could be the object of scholarly investigations. However, pragmatics has a long history of preferring – explicitly or implicitly – some types of data over other types, giving preference to unconstrained spoken interaction in natural settings. Written language, on the other hand, has often been rejected as unsuitable for pragmatic analyses because it is secondary (see section 3.1 below). Fictional language, as for instance in novels or plays, has met even more resistance because of its artificiality (see Jucker and Locher 2017). But even certain types of spoken language are occasionally seen as less ideal or unsuitable for pragmatic analyses, in particular language produced in highly constrained communicative situations, such as courtroom or classroom interaction because of the clear assignment of communicative roles and the constraints on the allowable contributions for the different participants.

Such an approach to language data that distinguishes between more acceptable and less acceptable types of data is based on an understanding of language as a more or less coherent and homogeneous entity where variations are seen as deviations. In such a framework, the linguist's task is considered to be the description of the common core of a language, and for this task only certain types of language use, such as maximally unconstrained spoken interaction, qualify as legitimate data. However, today many, perhaps most, pragmaticists adopt a very different view of language. Language is inherently variable and heterogeneous, and linguists are interested exactly in this variability. Every type of language has to be assessed on its own merits, and every type of language, whether spoken or written, deserves

to be investigated. This shift in perspective from homogeneity to heterogeneity has been identified as one of a number of paradigm shifts in linguistics (see Traugott 2008: 208; Jucker and Taavitsainen 2013: 6).

This handbook categorically adopts a perspective that focuses on the heterogeneity of language and on a diversity of research questions, data types, analytical approaches and theoretical frameworks. Its contributions provide overviews of a wide range of different methods of data collection and data analysis. In this first chapter, however, I shall provide an overview of the different types of data that are used by pragmaticists while the second chapter by Klaus P. Schneider provides an overview of the different types of data collection methods.

The early philosophers of language and pragmaticists, Austin and Searle, relied on their intuition in their seminal work on speech acts. Their data consisted of their own intuition about the use of language. As any competent native speaker, they knew what it meant to make a promise, to ask a question or to give advice, and they used this intuitive knowledge to dissect the relevant elements, what Austin (1962) called the felicity conditions, of these speech acts. In the philosophical tradition, intuition is an important source of data. According to Schneider,

> the word 'intuition' designates an uninferred or immediate kind of knowledge or apprehension, as opposed to discursive knowledge, mediated by accepted methods of demonstration. (Schneider 1995: 606)

For philosophers, it is important to discuss the possible foundation of such intuitive knowledge. "Introspection", according to Schneider, is a special type of intuitive knowledge. The objects of such knowledge are understood as being situated on an inner stage of a person. From there they can be retrieved by "looking inside" (Schneider 1995: 606). Feelings, emotions or the workings of our own native language are examples of such intuitive knowledge.

In the field of pragmatics, it is useful to make a terminological distinction between "intuition" or "intuitive knowledge" on the one hand and "introspection" on the other. Intuition here refers to the knowledge that a researcher brings to the task of investigating his or her native language, together with the ability to fabricate test sentences that can be assessed on the basis of their grammaticality or accessibility. The term "introspection", on the other hand, has been used for a long time in the fields of cognitive psychology and (applied) psycholinguistics to refer to experimental methods, involving thinking-aloud protocols and other elicitation techniques (see chapter 2 by Schneider). The papers in a volume edited by Færch and Kasper (1987), for instance, use the term introspection for a range of methods adopted from cognitive psychology, such as verbal reports by learners about their thought processes (see also Clark, this volume).

The terminological distinction helps to differentiate between the work of the language philosophers, who use their own intuitive knowledge about their native language to theorize about the use of language, and the work of experimental prag-

maticists, who use a range of elicitation techniques to access the native speaker introspection of the participants of their experiments. In a wider sense, all experimental work can be seen as accessing the introspection of native speakers. In production experiments, such as dialogue construction tasks or discourse completion tasks the elicited data consist of the language use that the participants consider to be typical or at least appropriate for a given situation. In comprehension and evaluation experiments, the introspective knowledge is accessed in a somewhat more direct form.

Before I turn to the different types of naturally occurring data, I will provide an outline of the different units of analysis in section 2. Section 3 will then be devoted to the different modalities of naturally occurring data and the different ways of conceptualizing these differences. It will cover not only the difference between spoken and written language but also the status of online and digital data, and the importance of sign language, i. e. the language systems used by deaf communities, and gestures as an additional layer of face-to-face communication. Section 4 deals with important dimensions or scales of observational data. It deals with constrained versus unconstrained language, and with the distinction between fictional and factual data. It also addresses the question of researcher interference. And it considers the difference between small snippets of data and huge corpora. This last dimension does not really concern the type of data under investigation but the research focus and whether the researcher attempts to discern communicative patterns on a micro scale of a short extract of a conversation, for instance, or whether the patterns are searched for across millions or even billions of words of running text.

2. Units of analysis

Utterances are – in a sense – the most basic units of analysis in pragmatics. They were the focus of the early language philosophers who asked how utterances can be used to change the world. Words are used to build utterances, which are used as speech acts to perform actions. Utterances are also the focus of researchers who are interested in how conversationalists interpret what they hear. Grice (1975), for instance, provides an account of how people systematically read between the lines of the utterances they hear; and Sperber and Wilson (1995) develop a comprehensive theory of utterance interpretation. Utterances are also seen as the main building blocks of larger structures, e. g. as turns-at-talk, where the focus is on the micro context of utterances and on the question of how one utterance is shaped by and helps to shape its immediate context. They are also seen as building blocks in layered hierarchies of conversational interactions (e. g. Sinclair and Coulthard 1975; Schiffrin 1987). In some cases, the focus of the pragmatic analysis is on units that are smaller than utterances, e. g. deictic elements, discourse markers,

stance markers and pragmatic noise. In other cases, it is on units that are larger than utterances, e. g. on entire discourses or texts or even on discourse domains. In this section, I would like to disentangle the different perspectives and give an overview of the units of analysis in pragmatics (see also Jucker 2008).

2.1. Utterances

The pioneering work of the language philosophers and early pragmaticists, Austin (1962) and Searle (1969), focused on what they called "speech acts", i. e. on utterances that are used to perform actions. Since this early work, the investigation of speech acts has been one of the most important pillars of pragmatic research. The early work relied on philosophical methods and the researcher's intuition about the nature of particular speech acts and how they are used to perform specific actions. Later work employed experimental methods, such as discourse completion tasks (e. g. Blum-Kulka, House and Kasper 1989), role plays and role enactments (e. g. Trosborg 1995), and, more recently, also corpus-linguistic methods (e. g. Deutschmann 2003). But in all cases the focus lies on single utterances and on how these utterances are used to perform specific actions. In some cases, the focus is extended to neighbouring speech acts. Compliments, for instance, regularly elicit responses, and some research, therefore, focuses on both elements of the pair and their sequential organisation (e. g. Golato 2005), but much of the research on compliments and compliment responses nevertheless focuses exclusively on either one or the other of the pair (see overview in Alfonzetti 2013).

Grice (1975) adopted a different perspective. He did not ask how utterances are used to perform actions but rather how conversationalists interpret utterances. How are listeners able to systematically read between the lines of what other people say? Utterances regularly mean more than what they explicitly say; they implicate additional meanings. Grice's Cooperative Principle is an attempt to give a systematic account of how listeners figure out the implicatures of individual utterances. Sperber and Wilson (1995), in their Relevance Theory, extended these questions to utterance interpretation in general. Listeners use pragmatic reasoning not only to recover implicatures from the utterances that they hear, but much more basically to work out even the explicitly communicated meaning of utterances. Blakemore's (1992) introductory textbook is even entitled *Understanding Utterances: An Introduction to Pragmatics*. Utterances, according to this theory, are underdetermined. They are ambiguous and vague. Nevertheless in actual situations, conversationalists generally pick out the intended meaning. They disambiguate and enrich the explicit content of utterances and come up with pragmatically meaningful interpretations of these utterances.

Utterances are also the building blocks of larger units. On a micro level, researchers have focused on the immediate context of utterances. It was the ethnomethodologists Sacks, Schegloff and Jefferson (1974), in particular, who initi-

ated a large body of research on the minutiae of the turn-taking system. They were interested in how one utterance – or turn-at-talk – is followed by another such unit with a minimal gap and no or minimal overlap between the units. This strand of research focuses on the transition between utterances and on the micro context in which utterances occur.

Researchers in this theoretical framework also noted that certain types of utterances tend to occur in pairs, so-called "adjacency pairs". Questions are followed by answers; greetings by greetings; invitations by acceptances or refusals and so on. This kind of research focuses on the pairings of utterances and on preferred or dispreferred combinations (see for instance Bilmes 1988; Schegloff 2007; Clift 2014). Dispreferred reactions, such as refusals or rejections, are generally clearly marked, while preferred reactions, such as acceptances, are generally unmarked. Thus, conversation analysis does not deal with utterance acts alone but with the sequencing of such acts, their interaction and the principles of their ordering.

With a slight shift of focus, utterances can also be seen as the building blocks of larger structures. Sinclair and Coulthard (1975), for instance, propose an analysis of classroom interaction consisting of a layered hierarchy (see e. g. Edmondson 2014). In this system, single utterances, and sometimes even parts of utterances are the smallest units, the so-called acts. They combine to form moves, such as "initiation", "response" or "feedback". Moves by different interlocutors combine to form exchanges. The three moves initiation, response and feedback, for instance, together form an exchange which is typical for classroom interaction. The teacher asks a question or uses some other way to initiate a reaction by the pupils. One of the pupils responds and the teacher gives feedback on the response. At a higher level, several exchanges combine to constitute transactions, which typically start with a preliminary exchange and – after a series of medial exchanges – end with a terminal exchange. Several transactions together, finally, make up an entire lesson.

2.2. Micro units (smaller than utterances)

While utterances may be seen as the most basic units of analysis in pragmatics, pragmaticists regularly focus on smaller elements as well. These elements have in common that their description requires pragmatic explanations, i. e. explanations that take into account the way in which these elements are used in actual situations and how they link the utterance in which they occur to the communicative situation in which they are used. Typical examples are deictic elements, stance markers, discourse markers, hedges and pragmatic noise. Deictic elements include a wide and diverse range of linguistic expressions which link the utterance in which they occur to its larger context (Levinson 1983: chapter 2; Chapman 2011: 39–42; Hanks 2011). Expressions like *now*, *then*, *next Thursday* or *this evening* connect the utterance in which they occur to its temporal frame; expressions like *here*, *on this side*, *behind*, *come* and *go* connect it to its spatial frame; and expressions like

but, *therefore*, *however*, *in conclusion* and *anyway* connect it to the discourse in which it occurs, to mention just the most important categories.

Stance markers are linguistic elements by which speakers convey their evaluations, personal attitudes and emotions as well as their level of commitment towards propositions (for an overview see Biber et al. 1999: chapter 12; Keisanen and Kärkkäinen 2014; Gray and Biber 2015; Landert 2017). They are a diverse group and – depending on the specific perspective – have been known under a variety of names, such as modality markers, subjectivity or intersubjectivity, hedges and so on. Typical linguistic elements that convey stance are modal and semi-modal verbs (e. g. *might* or *have to*), adverbials (adverbs, such as *obviously* or *fortunately*, or prepositional phrases, such as *in actual fact*) and complement clauses (e. g. *it's amazing that*). But stance can also be expressed through evaluative word choice and even with paralinguistic and non-linguistic means, including tone of voice, loudness, body posture, facial expression and gestures.

Discourse markers, too, comprise a heterogeneous set of elements that have received a lot of attention from pragmaticists with a range of different definitions and different terms. Schiffrin (1987: 31) defines them as "sequentially dependent elements which bracket units of talk", while Fraser (1999: 931) defines them as signalling "a relationship between the interpretation of the segment they introduce, S2, and the prior segment, S1". He further claims that "they have a core meaning, which is procedural, not conceptual, and their more specific interpretation is 'negotiated' by the context, both linguistic and conceptual" (Fraser 1999: 931; see Beeching 2016: chapter 1 for a discussion of different terms and definitions).

Pragmatic noise is a term that was introduced by Culpeper and Kytö (2010: chapters 9 to 12). They use it to refer "to items such as AH, HA, HAH, O, OH, HO, UM, HUM, as well as reduplicative forms like HA, HA or HA, HA, HA" (Culpeper and Kytö 2010: 199). They acknowledge that the term overlaps with the category of interjections, but it also includes laughter, pause-fillers and hesitation markers. Culpeper and Kytö investigate a corpus of Early Modern English dialogues, which means that they have to focus on the written representations of such elements in their data of plays and court records. But such elements have recently received more and more attention from researchers working on present-day materials (for an overview of work on pauses and hesitations see, for instance, Stenström 2011). Reber (2012) provides a detailed analysis of how speakers display affectivity in social interaction with a range of elements that she calls "sound objects", i. e. interjections, such as *oh*, *ooh* and *ah* or paralinguistic signals, such as whistles and clicks.

Thus pragmatic analyses regularly focus on linguistic elements that are smaller than utterances and indeed on paralinguistic and non-linguistic elements. But there is also a large body of work that focuses on entities larger than utterances.

2.3. Macro units (larger than utterances)

Utterances occur in contexts, and – as I have pointed out above – some researchers focus on the contextualisation of individual utterances into larger entities, be that as pairs of utterances or as entire discourses that are made up of structured sequences of utterances. Some pragmatic research, however, starts from a more global perspective and focuses primarily on larger units, which are variously called discourse or text. The term text is often restricted to written language while the term discourse is used for spoken language, but both terms are notoriously inconsistent across different research traditions (see Fetzer 2014 for an overview of different conceptualisations of the term discourse and Esser 2014 for an overview of taxonomies of discourse types).

A particular strand of this research goes back to the 1970s and 1980s and was originally labelled "textlinguistics". It was an attempt to seek linguistic regularities beyond the sentence boundaries, which manifested itself explicitly in book titles such as *A Text Grammar of English* (Werlich 1982). Werlich develops a typology of different types of text as well as an outline of the principles of text construction, their function and the contexts of their occurrence. In a similar way, de Beaugrande and Dressler (1981) investigate how texts are used in communication. Can we distinguish between acceptable and unacceptable texts in the same way that we can distinguish between grammatical and ungrammatical sentences? Which elements provide the cohesive ties that keep the sentences of a text together and render the text coherent? This tradition was particularly strong among German speaking scholars (see for instance the numerous introductions to textlinguistics written in German, e. g. Coseriu 1980; Sowinski 1983; Heinemann and Viehweger 1991; or more recently Schubert 2008). Many scholars tried to apply textlinguistic questions about the structure and function of texts to specific genres or text types. Suter (1993), for instance, focuses on wedding reports in local English newspapers. Wedding reports are descriptions of local weddings that have taken place in the week preceding the publication. The analysis focuses on the situational context in which these articles appear, the text production process, their content, thematic structure and their communicative function. Suter adds a diachronic dimension by contrasting wedding reports published in the 1930s to reports published in the 1980s. Auf dem Keller (2004) provides a similar analysis of textual structures in advertisements for books and medical supplies in eighteenth-century English newspapers. And Jacobs (2014) investigates press releases, i. e. texts from businesses, government agencies or political parties issued to the media in the hope of wider publicity.

The term discourse can not only be used to refer to the macro unit that is larger than the individual utterance, but also in a wider sense to refer to a discourse domain. In this sense, it refers to the entire range of linguistic practices in a particular, socially defined domain, as, for instance, in the discourse of sports or the discourse of science (Jucker 2008: 901; see also Henke 2005). Such domains are

large and overlapping. Historical pragmatics has a long tradition of investigating such domains, in particular, the discourse of science and mass media discourse (see, for instance, Jucker 2005; Claridge 2010; or the papers in Brownlees 2006 and Jucker 2009b).

In line with this chapter, this overview of units of analysis in pragmatics has focused on data units, such as utterances, discourse markers or entire discourses. Such a perspective does not cover the entire breadth of pragmatic research because pragmatic research does not always take a particular linguistic unit as a starting point. A good example would be the large area of politeness and impoliteness research. Here, the starting point is not a particular linguistic unit and how speakers use this unit in interaction, but rather particular types of interaction and the effects that such interaction has on the participants. This kind of research looks into the effects of communication and searches for elements that create these effects, whether they are words, such as terms of address, specific mitigators or speech acts (see e. g. Watts 2003). Cognitive approaches likewise do not take a linguistic unit as a starting point. They ask about the interrelationship between language and cognition (see Schmid 2012). Such approaches are interested in cognitive processes and how they are reflected in linguistic structures. They do not set out to analyse specific linguistic items, such as deictic elements, even if deictic elements may play a prominent role in their argumentation (see for instance Levinson 2003).

3. Medium of transmission

According to a simplistic view of language, there is a straightforward distinction between spoken language and written language. In one case we speak and listen, in the other we write and read. However, the situation is considerably more complex in particular for research in pragmatics. In the case of communication via electronic devices, such as computers, tablets and mobile phones the complexity increases even more. In addition to the spoken and the written mode, there is also sign language, which uses hand shapes and movements to communicate, and when we talk to each other, we also communicate with gestures, with our posture, with facial expressions and so on. The current section gives an overview of the important distinctions and introduces some of the models that have been developed to conceptualise them.

3.1. Spoken versus written language

The relationship between spoken and written language can be and has been described in many different ways. The written language, for instance, can be seen as derivative and secondary. By and large, all living languages have a spoken form but not all of them have a written form. Thus, the linguist's main task – one might

argue – involves the description of the spoken language. However, the advances of corpus linguistics have had the effect of shifting the primacy of description to the written language because language that already exists in written form is much more readily available for corpus compilation. Early corpora, such as the Brown Corpus or the LOB Corpus consisted entirely of written language, and even later corpora, such as the BNC or COCA only contain relatively small samples of transcribed spoken language. Biber et al.'s (1999) standard grammar of the English language treats the spoken language of conversation as a register next to fiction, news and academic texts. Biber (1988) investigated the variation between speech and writing in a systematic way. He contrasted large sections of the London-Lund corpus of spoken English and the LOB corpus of written English on the basis of features with specific discourse functions which he clustered into textual dimensions in order to evaluate specific texts according to their informational density, or their affective and interactional content.

It is probably fair to say that for a long time pragmaticists – in contrast to corpus linguists – ignored written language because of its secondary nature. However, there were also early attempts to think more carefully about the relationship between spoken and written language from a communicative or pragmatic point of view. Koch and Oesterreicher (1985, 2007; see also Koch 1999; Jucker and Taavitsainen 2013: 21–22), for instance, developed a model to clarify and visualise the distinction. They take the mode of transmission to be a dichotomy between phonic and graphic. Language is transmitted either in the phonic code or in the graphic code. In addition to this dichotomy, there is a scale between the opposite poles of communicative immediacy and communicative distance. The two codes are not restricted to one end of this scale but they have preferences. The graphic code has a preference for situations and genres of communicative distance while the phonic code has a preference for situations and genres of communicative immediacy. This is schematically illustrated in Figure 1.

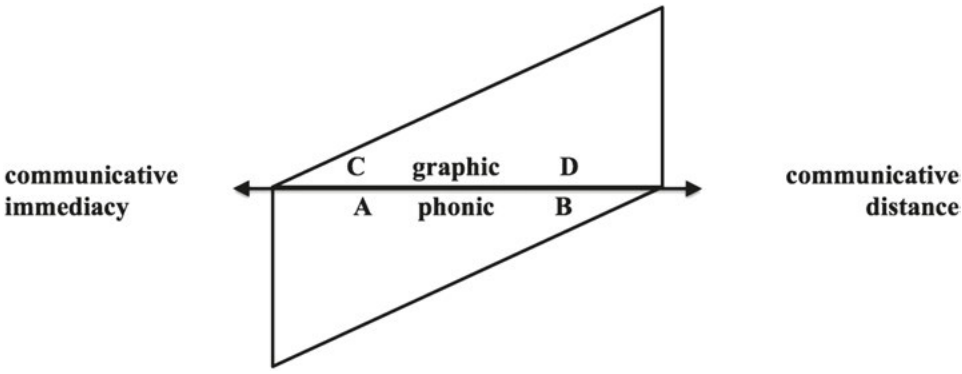

Figure 1: Koch and Oesterreicher's model of communicative immediacy and distance (Koch 1999: 400)

In this model, communicative immediacy is characterized by the parameters in the following list (Koch 1999: 400):

(a) physical (spatial, temporal) immediacy
(b) privacy
(c) familiarity of the partners
(d) high emotionality
(e) context embeddedness
(f) deictic immediacy (*ego-hic-nunc*, immediate situation)
(g) dialogue
(h) communicative cooperation of the partners
(i) free topic development
(j) spontaneity

Communicative distance, on the other hand, is characterized by the opposite values of these parameters, i. e. physical distance, lack of familiarity, low emotionality and so on. The four letters in the two triangles represent more typical or less typical situations. The letter A stands for communicative exchanges in the phonic code that are characterized by communicative immediacy, that is to say typically a face-to-face interaction between conversationalists who know each other well. The situation is informal, private, not public, and spontaneous. Topics can be freely chosen and changed and so on. But there are also communicative exchanges in the phonic code that are characterized by communicative distance; area B in the lower triangle. This applies to monologues, such as lectures, in formal, public situations with conversational partners who do not know each other well, and in situations that define specific topics and topic developments. Communicative immediacy is more typical for the phonic code. This is represented by the larger area A. Communicative distance is less typical, represented by the smaller area B. Most communicative situations in the phonic code are situated somewhere along the scale. Telephone conversations lack only a few of the communicative immediacy features of face-to-face conversations, while job interviews already show many of the communicative distance features of very formal, monologic situations.

The letters C and D stand for situations of communication in the graphic code. C represents the less typical situation of graphically communicated messages in situations of communicative immediacy. At the time when Koch and Oesterreicher developed their model this referred mainly to printed interviews, private letters, and entries in a personal diary. D represents the more typical situation of graphically communicated messages in situations of communicative distance. Legal texts, academic writing or articles in high-brow newspapers would be typical examples. Here, too, there are a lot of situations that are located between the two extremes. Today's situation with a wealth of typed messages transmitted electronically the situation has changed considerably. Communication via hand-held devices, via

social media and so on provide an entirely new situation for area C of communicative immediacy in the graphic code.

For researchers in historical pragmatics the relationship between the spoken and the written language is particularly important. In the early days of historical pragmatics, researchers often felt obliged to apologize for the use of written data in pragmatically driven investigations. In the absence of genuinely spoken data, they searched for instances of written language that were as close as possible to spoken language, such as dialogues in plays or transcripts of courtroom interactions. Rissanen (1986: 98), for instance, argued that "texts which record speech for some reason or another, are closer to spoken language than texts which are not based on actual speech". In fictional writing, the situation is even more complex. Authors regularly include oral features into their writings to give the dialogues of their characters an air of authenticity even if the features do not directly correspond to features attested in actual spoken discourse. Scripted and performed interactions between actors in plays also differ from normal everyday interactions in systematic ways (see Bublitz 2017).

In an early paper on historical pragmatics, Jucker (1998) sketched the various ways in which written language can be related to spoken language. Even genuinely written data can be classified into instances that tend to be monologic because there is normally little opportunity for the readers to interact with the writer and dialogic instances where such interaction is possible and expected.

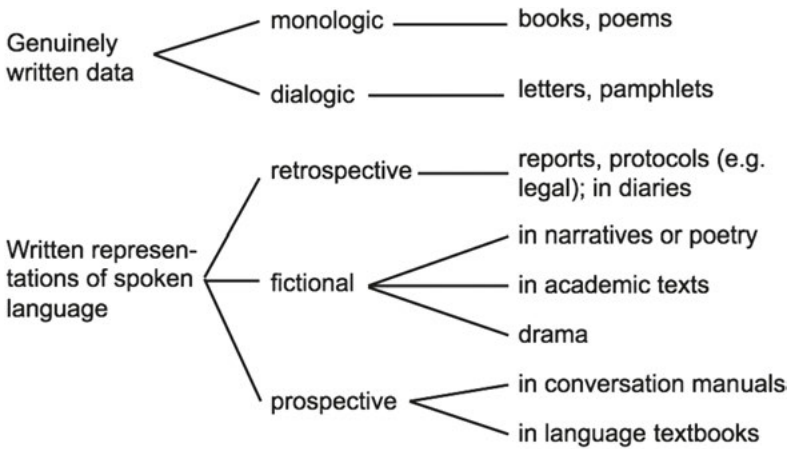

Figure 2: Data in historical pragmatics: the "communicative view" (Jucker 1998: 5, see also Jucker and Taavitsainen 2013: 23)

Written representations of spoken language can be separated into three different types. Reports, protocols and diaries regularly report actual spoken interactions, while conversation manuals and language textbooks record (invented) sample con-

versations that are meant to be used by the readers on future occasions. A third type is made up by fictional texts that record fictional conversations, for instance in play texts or in narrative literature but also, historically, in academic texts that were often written as fictional conversations between a master and a student (see also Culpeper and Kytö 2000 for a similar model and Kytö 2010 for an overview).

Landert and Jucker (2011) build on Koch and Oesterreicher in order to develop a model that adds two more dimensions to the dichotomy of phonic versus graphic and the scale of linguistic immediacy: the scale of accessibility and the scale of privacy. The distinction between phonic and graphic is not visually represented in their model. They argue that their model applies both to messages transmitted in the phonic code and to messages in the graphic code. In Figure 3, they provide prototypical examples from the sphere of graphically transmitted messages.

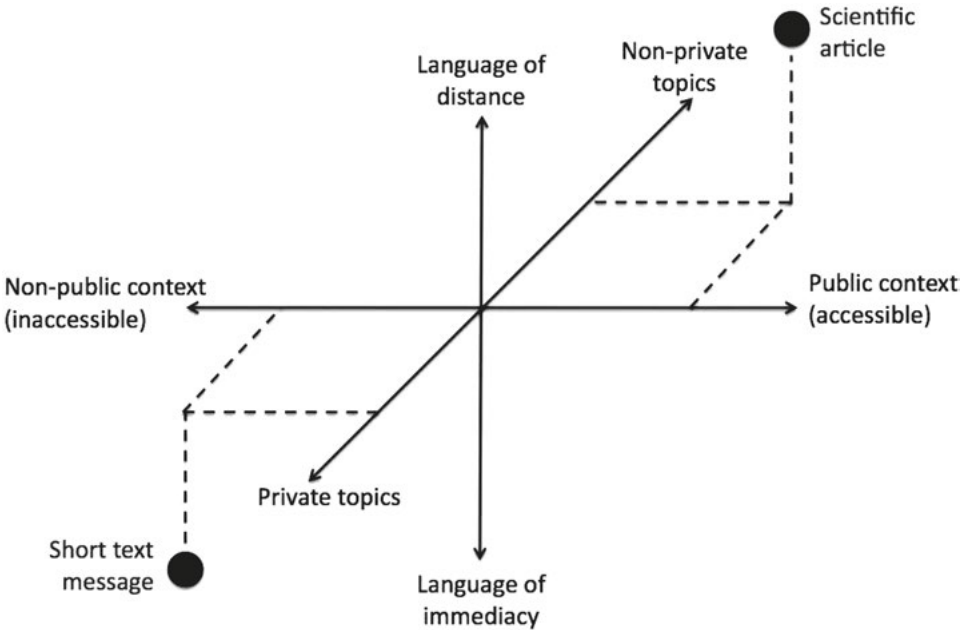

Figure 3: Enriched communicative model (Landert and Jucker 2011: 1427)

The scale of accessibility is defined by the ease of access to a particular message by others. In non-public situations, only very few people have access to a message. A typical message would be a short text message transmitted via a mobile phone intended for one single addressee. Such a message typically – but not necessarily – deals with private topics, which Landert and Jucker (2011: 1427) define as topics that "affect single individuals or very small groups of people", while non-private topics are topics "that lack this concentration on a private individual or a very small group". With this terminological move they disentangle the privacy of topics

from the accessibility of messages. This makes it possible, for instance, to describe more accurately what may be seen as a tendency in some sectors of today's mass media to make the private lives of celebrities public. The topics and issues remain private, according to this terminology, even if they are made public, i. e. publicly accessible. Scientific articles are prototypical examples of messages in the graphic code which deal with issues that are not restricted to a small group of individuals and which are made publicly accessible for a larger range of people.

3.2. Online/digital data

The communicative affordances of computer technology that have been developing over the last few decades have added new dimensions to the distinction between spoken and written language. To some extent there is still the dichotomy of the phonic code and the graphic code. We use computers and handheld devices to communicate with our voice (e. g. Internet telephony), and we use the same devices for all sorts of communication in the graphic code. However, the new technology has added an additional layer of affordances and has, therefore, opened up a large range of new research opportunities.

A clear terminology has not yet established itself for this type of communication. The most widespread term is probably "computer-mediated communication". It was already well-established in the 1990s and popularized by Herring (1996). There is a journal which uses this designation, the *Journal of Computer-Mediated Communication*, and a dedicated handbook (Herring, Stein and Virtanen 2013), entitled *Pragmatics of Computer-Mediated Communication*. But there are a host of other terms, such as "electronically mediated communication" or "electronic discourse", "digitally mediated communication" or "digital communication", "Internet-mediated communication", and "keyboard-to-screen communication" (see Crystal 2011: 1–3; Jucker and Dürscheid 2012: 35–37; or Locher 2014: 555–557 for a discussion of terminology). The different terms focus on different aspects of this special type of communication and they are not always entirely co-extensive in what they include or exclude. Herring's (2007: 1) definition of computer-mediated communication as "predominantly text-based human-human interaction mediated by networked computers or mobile telephone" explicitly includes communication via mobile phones, which begs the question whether mobile phones can be seen as computers. At the time when Herring proposed this definition, this was perhaps less clear than it is today. Terms such as "electronic discourse" (Locher 2014), "digital communication" (Tagg 2015) or "keyboard-to-screen communication" (Jucker and Dürscheid 2012) avoid the issue of classifying the electronic devices used to send and receive messages as computers or not and focus on the way in which the signals are transmitted or how they are encoded and received.

There are several important features that distinguish digital data from spoken and from written data. Spoken communication typically takes place in a situation

of synchronicity. The interactants are co-present, if not spatially then at least temporally (e. g. on the telephone). Messages are encoded and decoded at the same time. Written communication, on the other hand, typically takes place in an asynchronous situation. Messages are normally decoded only some time, perhaps even a very long time, after having been encoded. Computer-mediated communication uses the graphic code but it can be more or less synchronous. Jucker and Dürscheid (2012: 39) argue that the term "quasi-synchronous" is more appropriate for this type of communication. It covers all cases in which interactants exchange messages in quick succession, e. g. turn-by-turn in Facebook chat, or message-by-message in WhatsApp conversations. As such the term "quasi-synchronous" has fuzzy boundaries and coincides more or less with the term "synchronous" in cases where messages are transmitted not turn-by-turn, but stroke-by-stroke. And it coincides more or less with the term "asynchronous" in the case of, for instance, email messages that are exchanged in relatively quick succession.

Two further distinctions that are blurred in many forms of digital data are the oppositions between monologic and dialogic, and the opposition between discourse or text and utterance. Written communication tends to come in the form of monologic texts, while spoken communication most frequently comes in the form of dialogic utterances. For digital data, such a distinction is much less useful.

> For chat contributions, to take one specific example, neither the term "text" nor the term "utterance" seems to fit. They are realized in the graphic code, and thus may resemble a text. But they are also spontaneous, unplanned, context embedded (e. g. "What are you doing now?"), short and situated in a dialogic (more precisely: in a quasi-synchronous) context, and thus are more like prototypical utterances. (Jucker and Dürscheid 2012: 40)

As an alternative, Jucker and Dürscheid (2012: 42–44) propose the term "communicative act". Communicative acts can have a high expectation of being taken up and responded to by an interactant (in which case they are more utterance-like) or – at the other end of the scale – a small expectation of being taken up and responded to (in which case they are more text-like). Examples are chat contributions, which have a high expectation of uptake even if some contributions occasionally go unanswered, and user manuals, which have a very low expectation of uptake even if some frustrated user might occasionally try to get in contact with the author of the manual to complain about faulty or inscrutable instructions.

Digital data is further differentiated from traditional written data in its fluidity. Written texts, and in particular printed texts, are characterized by a high level of fixity. Once a text has been printed, it cannot easily be changed. Handwritten corrections within a printed text are easily recognizable as such. New printings of books are, of course, possible and common but each printing stays basically unalterable and fixed in its original form. This is not true for digital data. Texts that are stored digitally can easily be modified. Online news media, for instance, can update their texts on a minute-by-minute basis. This is why it has become standard

to add a time stamp to quotations of electronic texts. There is no guarantee that the text is still the same when it is checked some time later.

Finally, digital data are characterized by a vastly increased multimodality. Computer-mediated communication regularly combines language, images, memes, sounds and music. Still pictures and video clips have become very important in many forms of computer-mediated communication, especially on social-network sites or instant messaging applications, such as Facebook, Instagram, WhatsApp or Snapchat (see boyd 2014; Hoffmann and Bublitz 2017).

3.3. Sign language data

The term sign language is here used to refer to a class of languages used by deaf communities, such as German Sign Language or American Sign Language. They are equally complex in their structural features as spoken language, and, of course, they are not to be confused with the improvised gestures used by tourists in attempts to communicate with locals with whom they do not share a common language. In contrast to popular opinion, sign language is not only conveyed through hands but also through body language and facial expressions (Sutton-Spence and Woll 1999: 81; Quinn 2017: 55).

> Signs are, of course, a subset of human gestures, just as words are a subset of human vocalizations. Signs are distinguished from gestures by having an internal structure composed of elements which form a system of contrasts, and whose usage is rule-governed. (Woll and Kyle 1998: 855)

Like spoken language, sign language is ephemeral. If it is not recorded, it vanishes without a trace. Both spoken language and sign language are encoded and decoded at the same time, i. e. with synchronous production and reception. While the modality of spoken language is auditory, the modality of sign language is visual-spatial (Quinn 2017: 55). Relatively little is still known about the history of sign language in general and of specific sign languages. Recordings have only become available during the twentieth century. There are older accounts of deaf people who used signs to communicate (going back to Plato), but records or detailed descriptions of the signs that were used are missing (Woll and Kyle 1998: 855). One problem for the investigation of sign languages is that there is no generally accepted notation system. Moreover, photographs and drawings can only reproduce still pictures, and superimposed arrows can only provide a very limited rendering of the dynamics of signing and the way in which hand signs are accompanied and supported by body language and facial expressions (see, for instance, Sutton-Spence and Woll 1999: xi–xxi).

Pragmatic research on sign languages covers a wide spectrum. Groeber and Pochon-Berger (2014) as well as Cibulka (2016) deal with the peculiarities of turn-taking in signed conversations in Swiss German and in Swedish Sign Lan-

guage respectively. They focus on different types of holds, that is to say the freezing of a sign in turn-final position. The movement of the hand is momentarily suspended while hand shape and hand position are maintained. They show how holds perform important functions in the taking of turns and in the projectability of the next turn. Roush (2011), on the other hand, investigates issues of politeness and impoliteness in American Sign Language. In contrast to Groeber and Pochon-Berger (2014) and Cibulka (2016), who used a corpus of video recordings of signed interactions, he used an ethnographic approach by observing native signers in public gatherings of deaf communities and taking copious field notes (Roush 2011: 338). He focused in particular on metadiscursive terms and markers which were used to evaluate or describe the ongoing interaction. Mapson (2015) used data collected through semi-structured group discussions in order to analyse the ways in which professional interpreters developed their awareness of politeness in British Sign Language.

Kearsy, Smith and Zwets (2013) analysed the framing of constructed actions in British Sign Language narratives, and they used elicitation techniques in order to collect their data. 15 participants with British Sign Language as their preferred language were shown four short film clips and asked to retell the narratives to another deaf native signer of British Sign Language (one of the authors of the article) (Kearsy, Smith and Zwets 2013: 125).

3.4. Data of nonverbal behaviour

The importance of gestures and other forms of nonverbal behaviour in communication cannot be overestimated. As Kendon (2014) points out:

> Willingly or not, humans, when in co-presence, continuously inform one another about their intentions, interests, feelings and ideas by means of visible bodily action. For example, it is through the orientation of the body and, especially, through the orientation of the eyes, that information is provided about the direction and nature of a person's attention. (Kendon 2014: 1)

This opens up a vast range of research opportunities for pragmaticists, but there are various ways in which the scope of research can be focused on a subset of the visible bodily actions. The quotation above restricts the focus to those visible bodily actions that have an informative effect on a co-present human being, whether the effect was intended or not. The scope can be further reduced by restricting it to bodily actions that come with a communicative intention by the producer, that is to say actions that are meant to communicate. But this is a very fuzzy distinction and difficult to apply systematically. A more systematic restriction focuses on gestures that are used as part of an utterance, as for instance the use of hands in pointing to an object, in indicating the size or shape of an object or in emphasising what is being said. Cienki (2017) draws the line in a similar way. He focuses on

"movement of the hands and forearms by speakers when the movement is not part of an instrumental action (such as holding a pen and writing) and does not involve touching oneself or another (as in scratching one's head or patting someone on the back)" (2017: 61).

> We colour and flavour our speech with a variety of natural vocal, facial and bodily gestures, which indicate our internal state by conveying attitudes to the propositions we express or information about our emotions or feelings. Though we may be aware of them, such behaviours are often beyond our conscious control: they are involuntary or spontaneous. (Wharton 2009: 1)

Research of gestures and nonverbal behaviour shares some of the problems of research of sign languages. There is not a sufficiently established way of capturing the dynamic, spatio-temporal nature of gestures and other bodily actions in sufficient detail, but the problems are exacerbated for gestures because of the fuzzy nature of bodily actions that are relevant for communication (see Kendon 2014: Appendix 1 for a set of transcription conventions for gestural actions; see also Streeck 2009).

4. Observational data: Four dimensions

In the previous section, I focused on the different modalities of language and their relevance for pragmatic research. In this section, the focus shifts to four scalar dimensions that characterize observational data. The first dimension is the situational dimension, which distinguishes between speech contexts that are highly constrained in terms of what participants are expected – or indeed allowed – to say at specific points in the interaction and speech contexts that impose few – if any – such constraints on the contributions. The fictionality dimension distinguishes between fictional texts on the one hand and factual texts on the other. The third dimension distinguishes between different levels of researcher interference which ranges from data that came into existence without any researcher intervention and data that were purposefully elicited by a researcher. The fourth dimension, finally, distinguishes between researcher perspectives that focus on very small snippets of data to those that focus on a new generation of mega corpora. The first two dimensions are concerned with the nature of the data itself while the latter two are concerned with the researchers and their influence or perspective on the data.

All these dimensions are often invoked – explicitly or implicitly – in discussions about the suitability of certain types of data for specific research questions or even for pragmatic research in general. Here, they are not presented in an evaluative sense. There is no claim that one end of a particular scale is, in general, better than the other end, even though it may turn out to be better suited to specific types of research questions.

4.1. Situational dimension: Constrained versus unconstrained

Levinson (1979) defined the notion of "activity type" in terms of the allowable contributions and the constraints it imposes on participants, setting and so on:

> In particular I take the notion of an activity type to refer to a fuzzy category whose focal members are goal-defined, socially constituted, bounded, events with constraints on participants, setting, and so on, but above all on the kinds of allowable contributions. Paradigm examples would be teaching, a job interview, a jural interrogation, a football game, a task in a workshop, a dinner party and so on. (Levinson 1979: 368)

However, it seems clear that not all the activity types that he gives as examples are subjected to the same level of constraints. They can conveniently, but admittedly somewhat impressionistically, be situated on a scale from highly constrained situations to situations with relatively few constraints. At one end of the scale, we find speech situations that assign clear roles to the different participants and impose a large amount of restriction on the allowable contributions. Teaching, job interviews and jural interrogations are obvious examples. In each case the participants are assigned roles that come with very specific expectations as to the contributions that they are to make in this situation. Who asks questions? Who answers them? Who introduces new topics? And so on. At the other end of the scale we find speech situations in which there are no discernible role differences assigned by the situation. The dinner party mentioned by Levinson may be situated close to this end even though there are, of course, differences between the rights and obligations of the host or hostess and the guests. Other obvious examples might be a chat among friends on a long car drive, the locker room exchanges among the members of a sports team before or after a match, or the interactions of a group of children on the playground. In all these situations, there are also expectations as to what are appropriate or inappropriate contributions to the interaction, and some participants play a more important role while others play only subordinate roles. But the roles the individuals adopt are the result of the constellation of participants. They are not imposed by the speech situation in the way that an interview assigns differential roles to the interviewer and the interviewee.

Between the extreme cases there are interesting intermediate cases, such as a football game and a task in a workshop. A game of football imposes specific speaking rights to the referee, the coach and the team captain and imposes sanctionable restrictions on the allowable contributions by all the participants. But in contrast to interviews, spoken contributions are of subordinate importance, and there are few restrictions on the exchanges between the players themselves. A task in a workshop might also impose some restrictions on the allowable contributions, depending on the complexity of the task and the roles of the participants (e. g. supervisor and apprentice, etc.). An additional example would be a chat during a coffee break at a place of work. The situation itself may impose relatively few constraints but the

larger situation of the workplace with its differences in hierarchy may impose its own constraints on who initiates new topics and who breaks up the coffee break to go back to work.

The situational dimension is occasionally invoked in an evaluation of data in that unconstrained data is considered to be more genuine and, therefore, more likely to reveal the intricacies of conversational interaction without the interference of constraints imposed by the speech situation. However, the suitability of relatively constrained or relatively unconstrained data depends very much on the research question at hand. Speech situations, or activity types, cannot be placed on this scale with a high level of precision, but the scale itself helps to create an awareness for the varying importance of such constraints for specific situations.

4.2. Fictionality dimension: Fictional versus factual

Fictional language comes in many different guises. Obvious cases of fictional language are novels or short stories and other narratives that are the product of the imagination of an author without any claims to depict actual people and actual facts. It also includes theatre plays and telecinematic discourse, in which a scriptwriter invents dialogues that are performed by actors. But there is no clear-cut distinction between fictional data and non-fictional or factual data. Historical novels, for instance, may include depictions of historical figures next to invented figures within events that are partly historically attested and partly invented by the author. Television documentaries may include staged conversations performed by actors, and reality television may include a mixture of scripted and improvised conversations (see Jucker and Locher 2017: 5). Everyday conversations may include anecdotes, jokes and even personal narratives that consist of a mixture of factual and fictitious characters and events.

It is useful to draw a careful terminological distinction between the terms "fictional" and "fictitious" (see Klauk and Köppe 2014: 5–6; Jucker and Locher 2017: 6). The former refers to utterances, texts, pictures, movies, comics and so on, while the latter refers to characters, entities and events that have no correspondence outside of the text and do not exist in the real world. Fictional texts, then, deal with fictitious characters, entities and events. Factual texts, on the other hand, deal with characters, entities and events that have an existence in the real world, and in this sense texts can be factual even if they assert falsehoods about these characters, entities and events.

For a long time, pragmatics was not interested in fictional data. It was considered to be artificial, contrived and not sufficiently "real", and, therefore, not suitable for pragmatic analyses. Whenever pragmaticists, for instance in the area of historical pragmatics, resorted to fictional data, they felt the need to apologize for doing so (see for instance Brown and Gilman 1989: 159 or Salmon 1987: 265). They pointed out that in the absence of any "real" conversational data, fictional

data seemed to be a reasonably good approximation especially in the case of a skilful dramatist, such as William Shakespeare. Today, fictional data are seen as sufficiently interesting in themselves. They no longer serve as a substitute for "real" data but are analysed on their own terms. Many of Shakespeare's characters talk in iambic pentameters. It is safe to assume that at the turn from the sixteenth to the seventeenth century – or indeed at any other time – probably nobody used iambic pentameters in their everyday interactions. Shakespeare's dialogues do not represent real-life conversations but that does not make it less interesting to investigate the ways in which Shakespeare chose to depict his characters, how his characters interact, how they address each other, how they insult each other, how they are polite or impolite to each other and so on and so forth (see the collection of overviews of pragmatic approaches to fictional data in Locher and Jucker 2017).

4.3. Researcher interference dimension: Low versus high

The researcher interference scale relates to the amount of interference the researcher exerts on the production of language data. At one end of the scale there are language data that were produced entirely without the interference of a researcher. At the other end there are language data that were carefully elicited by a researcher in a highly controlled context. Figure 4 provides relevant examples along the scale.

Researcher Interference	Control		Relevant examples
Low	Low	1	Speech recording without researcher involvement
		2	Surreptitious recording by researcher
		3	Non-surreptitious recording by researcher
		4	Participant observation recording
		5	Semi-structured interview
		6	Role play or role enactment
		7	Dialog construction task
High	High	8	Oral DCTs

Figure 4: Researcher interference dimension

Speech recordings without any researcher involvement, number 1 in Figure 4, may, of course, be considered to be the most authentic type of data (Kasper 2000: 316) and, therefore, ideal for pragmatic research. It may be argued to be as close as possible to actual speech. However, with this type of recording the researcher depends entirely on the previous availability of data that were recorded for some non-research related purpose. Golato (2017) calls this "naturally occurring data" and refers to Potter's (2002: 541) "(conceptual) dead social scientist's test", which

asks whether the data would still exist even if the researcher got run over on the way to work. The researcher would not be able to carry out an interview, but a counselling session would take place even if the researcher failed to turn up.

Radio and television broadcasts are examples of recordings that do not depend on the presence of a researcher and – as forms of public spoken language – they are generally easily available. This makes them attractive as data for pragmaticists in spite of the lack of the researcher's control over the data. He or she cannot manipulate the situation in order to elicit special types of language patterns, e. g. specific speech acts and the like. The participants of such recordings are obviously aware of the fact that they are being recorded. The recording situation and a potentially very large audience are likely to constrain the language production of the participants in many ways. Thus, in spite of their usefulness, such recordings cannot be used as substitutes for unconstrained language use, and for many research questions such speech recordings are not available at all. Much of the content of the spoken component of corpora consist of such recordings. The *Corpus of Contemporary American English*, for instance, contains 109 million words of spoken language (out of a total of 520 million words), which consist entirely of transcripts of unscripted conversations from television and radio programmes (http://corpus.byu.edu/coca/) (see, for instance, Leech 2014: 256–260 on the inclusion of spoken language to corpora, such as the BNC or ICE).

This might make it interesting for researchers to collect the type of spoken data that they are interested in by setting up surreptitious recordings, number 2 in Figure 4. This would eliminate the observer's paradox (Labov 1972: 209) that we cannot observe behaviour when it is not being observed, but today's standards of ethical research – and in many countries even legal constraints – rule out such a procedure (see, for instance, Duranti 1997: 117; Flöck 2016: 36). It is no longer acceptable – as apparently it was in the early days of speech recordings – to record people surreptitiously and only ask them after the event (but see Hambling-Jones and Merrison's 2012: 1121 argumentation that in some situations surreptitious recordings and retrospective consent might be superior to pre-obtained consent).

With non-surreptitious recordings, number 3 in Figure 4, the researcher has to accept the observer's paradox and the effects that the recording equipment might have on the participants. This category, of course, comprises a rather large range of possible situations from dinner table conversations to specifically elicited narratives or service encounter recordings. In some cases, the researcher takes part in the conversations that he or she records, which turns them into participant observation recordings. Schiffrin (1987), for instance, carried out what she called sociolinguistic interviews with groups of people from her neighbourhood and with whom she shared an ethnic identity. She points out how her participation complicated the observer's paradox (Schiffrin 1987: 41). The analyst's role might influence the development of the interaction and it might influence the interpretation of the results because the analyst is no longer a neutral outsider. Rüegg (2014), to

mention a more recent example, investigated thanks responses from a variational perspective. She collected her data by recording visits to restaurants in Los Angeles in three different price ranges. The recordings of the interactions between a waiter and a small group of guests were not surreptitious but the interactions clearly had a primary purpose that was outside the linguistic research questions. They had to do with offering and ordering food and drinks and with the incidental necessities of serving food and drinks, clearing the table and so on.

Number 1 to 4 on the researcher interference dimension can all still be considered "naturally occurring data" but it is clear that there are differences in the level of researcher interference and – concomitantly – in the level of researcher control. With participant observations, the researcher can, of course, try to influence the flow of the conversation and thus take at least some control of what kind of language the participants produce, especially if they manage to create speech situations in which the pragmatic element under investigation is likely to occur in a naturalistic way because of the necessities imposed by the situation.

The remaining numbers on the dimension shift the balance from naturally occurring data to elicited data (dealt with in more detail in Schneider, this volume). They impose more and more control on the language production of the participants. While a semi-structured ethnographic interview, number 5, leaves some room for a broader range of responses from participants, role plays or role enactment tasks, number 6, ask for very specific behaviour, in which the responses depend – at least to some extent – on the acting abilities of the participants and their willingness to play along. Dialog construction tasks, number 7, ask participants to create – usually in written form – an entire dialogue including the utterances by several participants in order to elicit the participants' intuition about typical or appropriate dialogues in a given situation. Discourse completion tasks, number 8, finally impose the highest level of control on the participants' language production. Usually they are expected to produce a speech act of a very specific type, such as an apology, a request or a response to a compliment.

4.4. Researcher perspective dimension: Micro versus macro

The researcher perspective dimension relates to the amount of data that is being investigated. It does not distinguish between different types of data as the three dimensions outlined above. It is concerned with the perspective adopted by the researcher. At one end of the scale the researcher investigates a very small amount of usually richly contextualized data, prototypically a single conversation or even just a small extract of a conversation where the researcher knows a lot about the participants and the context in which the conversation took place. At the other end of the scale the researcher searches for patterns of language use in large corpora consisting of millions or even billions of words. Bednarek (2011: 546) illustrates this dimension with Figure 5:

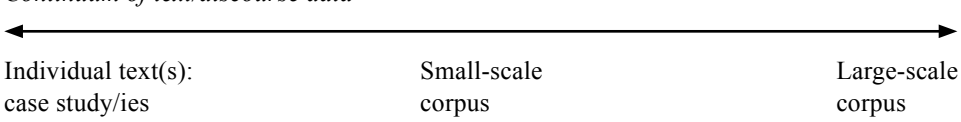

Figure 5: Researcher perspective dimension (Bednarek 2011: 546)

Case studies of individual texts allow for rich contextualisations while large-scale corpora only provide very minimal contextualisations, that is to say the amount of data and the contextual richness can be seen – in a very abstract way – as a reciprocal function. With an increasing amount of data, the contextual richness becomes smaller and smaller. And, reciprocally, high contextual richness can only be achieved if the amount of data is very small. The investigative precision, to use Leech's term with a slightly different meaning, does not favour one over the other. In fact, the investigative precision can only be increased by increasing the amount of data with a given value of contextual richness or vice versa.

A brief example may illustrate these interdependencies. The phrase *I'm sorry* generally serves as an apology, whose occurrences can be investigated both in a small-scale case study or in a large-scale corpus. In Barbara Kingsolver's novel *Flight Behavior* (2012), there are 20 instances of *I'm sorry*. Each and every one of these instances is richly contextualized, and the reader can work out the level of sincerity that is attached to each one, whether it is a token apology for an interruption as in (1) or whether it is a heart-felt apology for breaking up a marriage as in (2). In an important sense, fictional examples provide a more complete contextualization than real life conversations. In real life, conversationalists under observation from a researcher have a wealth of life experiences that are not accessible to the researcher. In a novel, the depicted characters do not have any life experiences outside of the novel. Whatever is relevant for the novel is depicted in the novel.

(1) "I'm sorry for the interruption, Bobby," Brenda's mother said, cocking one hand on her hip, doing a poor job of looking sorry. (Page 99, Location 1198)
(2) "I'm sorry," she said. "I'm thankful for our children. But I'm not what you need." (Page 527, Location 6500)

Figure 7 shows the frequency development of *I'm sorry* over two centuries of American English. It is based on a corpus of digitized texts containing more than five million books and a total of some 361 billion words in English texts (Michel et al. 2010). But these instances are entirely decontextualized. For copy-right reasons the software does not access a database containing all these books but indexed lists of ngrams derived from these books. Each ngram in the database comes with an indication of the year of publication and its language or language variety but it is

disconnected from its actual context. This is an extreme case of a decontextualized database, and it is, therefore, usually shunned by corpus linguists except for some very preliminary initial searches that can be used to ask more specific questions. In this case, it is impossible to ascertain, for instance, whether the phrase *I'm sorry* was indeed used as an apology or perhaps to perform another speech act, as for instance the expression of condolences.

Figure 6 shows that the phrase had a very low frequency in the nineteenth century. Its use increased in the first half of the twentieth century with a noticeable decrease in the 1960s and 1970s and a sharp increase after that, which poses interesting follow-up questions whether the decrease in the 1960s and 1970s could in any way be related to social and cultural developments at the time. However, in order to tackle such questions, the research would have to go back to contextualized data samples (see also O'Keeffe, this volume on the development of *I'm sorry* versus *I apologise*).

5. Conclusion

Pragmatics studies the use of language in all its complexities and diversities, which means that language in all its various forms, shapes and varieties provides the data for pragmatic research. Pragmatics no longer focuses on a single type of data, such as, for instance, spontaneous, multi-party conversations that take place in private settings. Pragmatics is not restricted to the modality of spoken language. It is also concerned with written language, with digital language, with sign language and with all aspects of nonverbal communication. Different types of language data invite different types of research questions, and different research questions require different types of data, as well as different methods of collecting and analysing it (see Félix-Brasdefer 2007; Jucker 2009a; Golato 2017).

In many cases, it is the triangulation of different types of data that provide a better understanding of pragmatic issues. Félix-Brasdefer (2007: 163), for instance, uses both role play data and naturally occurring interactions in his study of requests in Mexican Spanish, and Flöck (2016: 84), who compares requests in British English and in American English, uses both audio recordings of informal, naturally occurring conversations and written production data elicited in discourse completion tasks. In both studies the combination of data and methods provided a more comprehensive view of requests than a reliance on one type of data would have made possible.

This introductory chapter has given an overview of different types of data in pragmatic research (data collection methods are covered by Schneider, this volume). Such a task is potentially boundless because virtually all the existing literature in pragmatic research could be situated within the scope of this paper. I have, therefore, focused on the relevant modalities (spoken, written, digital, signed,

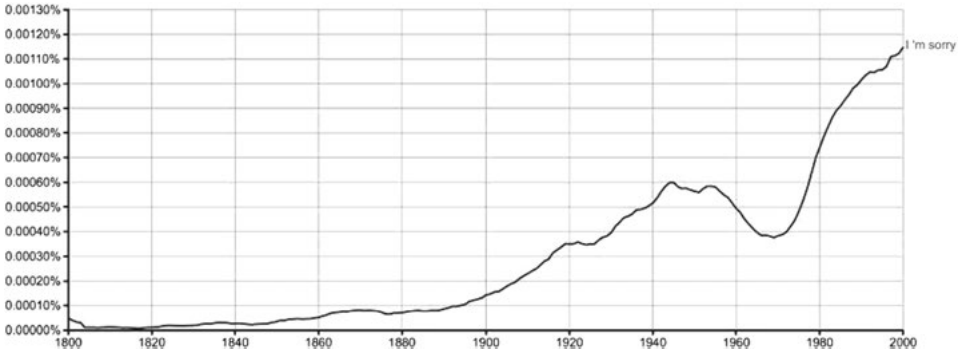

Figure 6: "I'm sorry" in American English from 1800 to 2000 (http://books.google.com/ngrams/)

nonverbal) and their impact on pragmatic research as well as the relevant data dimensions (level of constraints and fictionality) and researcher dimensions (interference/control and research perspective/data size).

Acknowledgment

A special word of thanks goes to Andrea Golato, Daniela Landert, Magdalena Leitner, Miriam Locher, Mirjam Schmalz and Larssyn Staley as well as to my two co-editors of this volume for their helpful and perceptive comments on draft versions of this chapter. The usual disclaimers apply.

Data sources

Kingsolver, Barbara
2012 *Flight Behavior. A Novel*. New York: Harper (Kindle Edition).

References

Alfonzetti, Giovanna
2013 Compliments. In: Marina Sbisà and Ken Turner (eds.), *Pragmatics of Speech Actions*, 555–586. (Handbooks of Pragmatics 2.) Berlin: de Gruyter Mouton.
Auf dem Keller, Caren
2004 *Textual Structures in Eighteenth-century Newspaper Advertising. A Corpus-based Study of Medical Advertisements and Book Advertisements*. Aachen: Shaker.

Austin, John L.
1962 *How to Do Things with Words. The William James Lectures Delivered at Harvard University in* 1955. Oxford: Oxford University Press.
Beaugrande, Robert-Alain de and Wolfgang Ulrich Dressler
1981 *Introduction to Text Linguistics*. London: Routledge.
Bednarek, Monika
2011 Approaching the data of pragmatics. In: Wolfram Bublitz and Neal R. Norrick (eds.), *Foundations of Pragmatics*, 537–559. (Handbooks of Pragmatics 1.) Berlin: de Gruyter Mouton.
Beeching, Kate
2016 *Pragmatic Markers in British English. Meaning in Social Interaction*. Cambridge: Cambridge University Press.
Biber, Douglas
1988 *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan
1999 *Longman Grammar of Spoken and Written English*. London: Longman.
Bilmes, Jack
1988 The concept of preference in conversation analysis. *Language in Society* 17(2): 161–181.
Blakemore, Diane
1992 *Understanding Utterances. An Introduction to Pragmatics*. Oxford: Blackwell.
Blum-Kulka, Shoshana, Juliane House and Gabriele Kasper (eds.)
1989 *Cross-Cultural Pragmatics: Requests and Apologie*s. Norwood, NJ: Ablex.
boyd, danah
2014 *It's Complicated: The Social Lives of Networked Teens*. New Haven and London: Yale University Press.
Brown, Roger and Albert Gilman
1989 Politeness theory and Shakespeare's four major tragedies. *Language in Society* 18(2): 159–212.
Brownlees, Nicholas (ed.)
2006 *News Discourse in Early Modern Britain. Selected Papers of CHINED* 2004. Bern: Peter Lang.
Bublitz, Wolfram
2017 Oral features in fiction. In: Miriam A. Locher and Andreas H. Jucker (eds.), *Pragmatics of Fiction*, 235–263. (Handbooks of Pragmatics 12.) Berlin: de Gruyter Mouton.
Chapman, Siobhan
2011 *Pragmatics*. (Palgrave Modern Linguistics.) Houndmills, Basingstoke: Palgrave Macmillan.
Cibulka, Paul
2016 On how to do things with holds. Manual movement phases as parts of interactional practices in signed conversation. *Sign Language Studies* 16(4): 447–472.
Cienki, Alan
2017 Gesture and pragmatics: From paralinguistic to variably linguistic. In: Anne

Barron, Peter Grundy and Gu Yueguo (eds.), *Routledge Handbook of Pragmatics*, 61–68. London: Routledge.

Claridge, Claudia
2010 News discourse. In: Andreas H. Jucker and Irma Taavitsainen (eds.), *Historical Pragmatics*, 587–620. (Handbooks of Pragmatics 8.) Berlin: de Gruyter Mouton.

Clift, Rebecca
2014 Conversation analysis. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 97–124. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.

Coseriu, Eugenio
1980 *Textlinguistik. Eine Einführung*. (Tübinger Beiträge zur Linguistik 109.) Tübingen: Gunter Narr.

Crystal, David
2011 *Internet Linguistics: A Student Guide*. London: Routledge.

Culpeper, Jonathan and Merja Kytö
2000 Data in historical pragmatics: Spoken discourse (re)cast as writing. *Journal of Historical Pragmatics* 1(2): 175–199.

Culpeper, Jonathan and Merja Kytö
2010 *Early Modern English Dialogues. Spoken Interaction as Writing*. (Studies in English Language.) Cambridge: Cambridge University Press.

Deutschmann, Mats
2003 *Apologising in British English*. (Skrifter från moderna språk 10.) Umeå: Institutionen för moderna språk, Umeå University.

Duranti, Alessandro
1997 *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Edmondson, Willis J.
2014 The emergence of discourse analysis as a disciplinary field: philosophical, pedagogic and linguistic approaches. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 65–96. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.

Esser, Jürgen
2014 Taxonomies of discourse types. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 443–462. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.

Færch, Claus and Gabriele Kasper (eds.)
1987 *Introspection in Second Language Research*. Clevedon: Multilingual Matters.

Félix-Brasdefer, J. César
2007 Natural speech vs. elicited data. *Spanish in Context* 4(2): 159–185.

Fetzer, Anita
2014 Conceptualising discourse. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 35–62. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.

Flöck, Ilka
2016 *Requests in American and British English*. (Pragmatics & Beyond New Series 265.) Amsterdam: John Benjamins.

Fraser, Bruce
1999 What are discourse markers? *Journal of Pragmatics* 31(7): 931–952.
Golato, Andrea
2005 *Compliments and Compliment Responses: Grammatical Structure and Sequential Organization*. (Studies in Discourse and Grammar 15.) Amsterdam: John Benjamins.
Golato, Andrea
2017 Naturally occurring data. In: Anne Barron, Yuego Gu and Gerard Steen (eds.), *The Routledge Handbook of Pragmatics*, 21–26. London: Routledge.
Gray, Bethany and Douglas Biber
2015 Stance markers. In: Karin Aijmer and Christoph Rühlemann (eds.), *Corpus Pragmatics. A Handbook*, 219–248. Cambridge: Cambridge University Press.
Grice, Paul H.
1975 Logic and conversation. In: Peter Cole and J.L. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press.
Groeber, Simone and Evelyne Pochon-Berger
2014 Turns and turn-taking in sign language interaction: A study of turn-final holds. *Journal of Pragmatics* 65: 121–136.
Hambling-Jones, Oliver and Andrew John Merrison
2012 Inequity in the pursuit of intimacy: An analysis of British pick-up artist interactions. *Journal of Pragmatics*. Special Issue: *Im/politeness across Englishes* 44(9): 1115–1127.
Hanks, William F.
2011 Deixis and indexicality. In: Wolfram Bublitz and Neal R. Norrick (eds.). *Foundations of Pragmatics*, 315–346. (Handbooks of Pragmatics 1.) Berlin: de Gruyter Mouton.
Heinemann, Wolfgang and Dieter Viehweger
1991 *Textlinguistik. Eine Einführung*. (Germanistische Linguistik 115.) Tübingen: Niemeyer.
Henke, Christoph
2005 Diskursanalyse und Literatur: Michel Foucaults Anti-Hermeneutik. In: Hans Vilmar Geppert und Hubert Zapf (eds.), *Theorien der Literatur*, Vol. 2, 243–260. Würzburg: Königshausen & Neumann.
Herring, Susan C.
2007 A faceted classification scheme for computer-mediated discourse. language@internet 1/2007 (http://www.languageatinternet.de)
Herring, Susan C. (ed.)
1996 *Computer-Mediated Communication. Linguistic, Social and Cross-Cultural Perspectives*. (Pragmatics & Beyond New Series 39.) Amsterdam: John Benjamins.
Herring, Susan, Dieter Stein and Tuija Virtanen (eds.)
2013 *Pragmatics of Computer-Mediated Communication*. (Handbooks of Pragmatics 9.) Berlin: de Gruyter Mouton.
Hoffmann, Christian and Wolfram Bublitz (eds.)
2017 *Pragmatics of Social Media*. (Handbooks of Pragmatics 11.) Berlin: de Gruyter Mouton.

Jacobs, Geert
2014 Press releases. In: Klaus P. Schneider, and Anne Barron (eds.), *Pragmatics of Discourse*, 583–600. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.
Jucker, Andreas H.
1998 Historical pragmatics: An interdisciplinary approach. In: Raimund Borgmeier, Herbert Grabes and Andreas H. Jucker (eds.), *Anglistentag 1997 Giessen. Proceedings*, 3–7. Trier: Wissenschaftlicher Verlag.
Jucker, Andreas H.
2005 News discourse: Mass media communication from the seventeenth to the twenty-first century. In: Janne Skaffari, Matti Peikola, Ruth Carroll, Risto Hiltunen and Brita Wårvik (eds.), *Opening Windows on Texts and Discourses of the Past*, 7–21. (Pragmatics & Beyond New Series 134.) Amsterdam: John Benjamins.
Jucker, Andreas H.
2008 Historical pragmatics. *Language and Linguistics Compass* 2(5): 894–906.
Jucker, Andreas H.
2009a Speech act research between armchair, field and laboratory: The case of compliments. *Journal of Pragmatics* 41(8): 1611–1635.
Jucker, Andreas H. (ed.)
2009b *Early Modern English News Discourse. Newspapers, Pamphlets and Scientific News Discourse*. (Pragmatics & Beyond New Series 187.) Amsterdam: John Benjamins.
Jucker, Andreas H. and Christa Dürscheid
2012 The linguistics of keyboard-to-screen communication: A new terminological framework. *Linguistik online* 56, 6/12. http://www.linguistik-online.org/56_12/
Jucker, Andreas H. and Miriam A. Locher
2017 Introducing *Pragmatics of Fiction*: Approaches, trends and developments. In: Miriam A. Locher and Andreas H. Jucker (eds.), *Pragmatics of Fiction*, 1–21. (Handbooks of Pragmatics 12.) Berlin: de Gruyter Mouton.
Jucker, Andreas H. and Irma Taavitsainen
2013 *English Historical Pragmatics*. (Edinburgh Textbooks on the English Language.) Edinburgh: Edinburgh University Press.
Kasper, Gabriele and Merete Dahl
1991 Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition* 13(2): 215–247.
Kasper, Gabriele
2000 Data collection in pragmatics research. In: Helen Spencer-Oatey (ed.), *Culturally Speaking. Managing Rapport Through Talk Across Cultures*, 316–341. London: Continuum.
Kearsy, Cormier, Sandra Smith and Martine Zwets
2013 Framing constructed action in British Sign Language narratives. *Journal of Pragmatics* 55: 119–139.
Keisanen, Tiina and Elise Kärkkäinen
2014 Stance. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 295–322. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.

Kendon, Adam
2014 *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.
Klauk, Tobias und Tilmann Köppe
2014 Bausteine einer Theorie der Fiktionalität. In: Tobias Klauk und Tilmann Köppe (Hrsg.), *Fiktionalität. Ein interdisziplinäres Handbuch*, 3–31. Berlin: de Gruyter.
Koch, Peter
1999 Court records and cartoons: Reflections of spontaneous dialogue in Early Romance texts. In: Andreas H. Jucker, Gerd Fritz and Franz Lebsanft (eds.), *Historical Dialogue Analysis*, 399–429. (Pragmatics & Beyond New Series 66.) Amsterdam: John Benjamins.
Koch, Peter und Wulf Oesterreicher
1985 Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36: 15–43.
Koch, Peter und Wulf Oesterreicher
2007 Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik. Deutsche Sprache in Gegenwart und Geschichte* 35(3): 346–375.
Kytö, Merja
2010 Data in historical pragmatics. In: Andreas H. Jucker and Irma Taavitsainen (eds.), *Historical Pragmatics*, 33–67. (Handbooks of Pragmatics 8.) Berlin: de Gruyter Mouton.
Labov, William
1972 *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
Landert, Daniela
2017 Stance in fiction. In: Miriam A. Locher and Andreas H. Jucker (eds.), *Pragmatics of Fiction*, 489–514. (Handbooks of Pragmatics 12.) Berlin: de Gruyter Mouton.
Landert, Daniela and Andreas H. Jucker
2011 Private and public in mass media communication. From letters to the editor to online commentaries. *Journal of Pragmatics* 43: 1422–1434.
Leech, Geoffrey N.
2014 *The Pragmatics of Politeness*. (Oxford Studies in Sociolinguistics.) Oxford: Oxford University Press.
Levinson, Stephen C.
1979 Activity types and language. *Linguistics* 17: 365–399.
Levinson, Stephen C.
1983 *Pragmatics*. Cambridge: Cambridge University Press.
Levinson, Stephen C.
2003 *Space in Language and Cognition. Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.
Locher, Miriam A.
2014 Electronic discourse. In: Klaus P. Schneider and Anne Barron (eds.), *Pragmatics of Discourse*, 555–581. (Handbooks of Pragmatics 3.) Berlin: de Gruyter Mouton.
Locher, Miriam A. and Andreas H. Jucker (eds.)
2017 *Pragmatics of Fiction*. (Handbooks of Pragmatics 12.) Berlin: de Gruyter Mouton.

Mapson, Rachel
2015 Paths to politeness: Exploring how professional interpreters develop an understanding of politeness norms in British Sign Language and English. In: Barbara Pizziconi and Miriam A. Locher (eds.), *Teaching and Learning (Im)politeness*, 155–184. Berlin: de Gruyter Mouton.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak and Erez Lieberman Aiden
2010 Quantitative analysis of culture using millions of digitized books. *Science* (http://www.sciencemag.org/content/early/2010/12/15/science.1199644)

Potter, Jonathan
2002 Two kinds of natural. *Discourse Studies* 4(4): 539–542.

Quinn, Gary
2017 British Sign Language. In: Anne Barron, Peter Grundy and Gu Yueguo (eds.), *Routledge Handbook of Pragmatics*, 55–60. London: Routledge.

Reber, Elisabeth
2012 *Affectivity in Interaction. Sound Objects in English*. (Pragmatics & Beyond New Series 215.) Amsterdam: John Benjamins.

Rissanen, Matti
1986 Variation and the study of English historical syntax. In: David Sankoff (ed.), *Diversity and Diachrony*, 97–109. Amsterdam: John Benjamins.

Roush, Daniel R.
2011 Language between bodies: A cognitive approach to understanding linguistic politeness in American Sign Language. *Sign Language Studies* 11(3): 329–374.

Rüegg, Larssyn
2014 Thanks responses in three socio-economic settings: A variational pragmatics approach. *Journal of Pragmatics* 71: 17–30.

Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson
1974 A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696–735.

Salmon, Vivian
1987 Sentence structures in colloquial Shakespearian English. In: Vivian Salmon and Edwina Burness (eds.), *A Reader in the Language of Shakespearean Drama*, 265–300. Amsterdam: John Benjamins. First published in *Transactions of the Philological Society* 64(1): 105–140 [1965].

Schegloff, Emanuel A.
2007 *Sequence Organization in Interaction. A Primer in Conversation Analysis I*. Cambridge: Cambridge University Press.

Schiffrin, Deborah
1987 *Discourse Markers*. (Studies in Interactional Sociolinguistics 5.) Cambridge: Cambridge University Press.

Schmid, Hans-Jörg (ed.)
2012 *Cognitive Pragmatics*. (Handbooks of Pragmatics 4.) Berlin: de Gruyter Mouton.

Schneider, Hans Julius
1995 Intuition and introspection. In: Jef Verschueren, Jan-Ola Östman and Jan

Blommaert (eds.), *Handbook of Pragmatics. Manual*, 606–608. Amsterdam: John Benjamins.

Schubert, Christoph
2008 *Englische Textlinguistik. Eine Einführung*. (Grundlagen der Anglistik und Amerikanistik 30.) Berlin: Erich Schmidt.

Searle, John R.
1969 *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sinclair, John McH. and R. Malcolm Coulthard
1975 *Towards an Analysis of Discourse. The English Used by Teachers and Pupils*. Oxford: Oxford University Press.

Sowinski, Bernhard
1983 *Textlinguistik. Eine Einführung*. Stuttgart: Kohlhammer.

Sperber, Dan and Deirdre Wilson
1995 *Relevance. Communication and Cognition*. Second edition. Oxford: Basil Blackwell.

Stenström, Anna-Brita
2011 Pauses and hesitations. In: Gisle Andersen and Karin Aijmer (eds.), *Pragmatics of Society*, 537–567. (Handbooks of Pragmatics 5.) Berlin: de Gruyter Mouton.

Streeck, Jürgen
2009 *Gesturecraft. The Manu-facture of Meaning*. (Gesture Studies 2.) Amsterdam: John Benjamins.

Suter, Hans-Jürg
1993 *The Wedding Report. A Prototypical Approach to the Study of Traditional Text Types*. (Pragmatics & Beyond New Series 27.) Amsterdam: Benjamins.

Sutton-Spence, Rachel and Bencie Woll
1999 *The Linguistics of British Sign Language. An Introduction*. Cambridge: Cambridge University Press.

Tagg, Caroline
2015 *Exploring Digital Communication. Language in Action*. (Routledge Introductions to Applied Linguistics.) London: Routledge.

Traugott, Elizabeth Closs
2008 The state of English language studies: A linguistic perspective. In: Marianne Thormählen (ed.), *English Now. Selected Papers from the 20th IAUPE Conference in Lund 2007*, 199–225. Lund: Lund Studies in English.

Trosborg, Anna
1995 *Interlanguage Pragmatics. Requests, Complaints and Apologies*. Berlin: Mouton de Gruyter.

Watts, Richard J.
2003 *Politeness*. (Key Topics in Sociolinguistics.) Cambridge: Cambridge University Press.

Werlich, Egon
1982 *A Text Grammar of English*. (UTB 597.) Second edition. Heidelberg: Quelle & Meyer.

Wharton, Tim
2009 *Pragmatics and Non-Verbal Communication*. Cambridge: Cambridge University Press.

Woll, Bencie and Jim G. Kyle
1998 Sign language. In: Jacob L. Mey (ed.), *Concise Encyclopedia of Pragmatics*, 854–878. Amsterdam: Elsevier.